
From: Gomez, Enrique [Enrique.Gomez@dps.texas.gov]
Sent: Friday, December 09, 2011 4:38 PM
To: Gomez, Enrique; Lee Guyette; Moon, Charles; John Mendoza
Cc: O'Connor, Ryan; Johnson, Lance; Martinez, Germaine
Subject: RE: jury wheel ftp address

Lee,
My team tell me that you guys already received the file.

Enrique

From: Gomez, Enrique
Sent: Friday, December 09, 2011 4:33 PM
To: 'Lee Guyette'; Moon, Charles; John Mendoza
Cc: O'Connor, Ryan; Johnson, Lance; Martinez, Germaine
Subject: RE: jury wheel ftp address

Lee,
We'll be ready to send you the data on Monday.
These are the filters that Charles used:

1. Select first, middle, last name, or City not containing "TEST"
2. Select first, middle, or last name not containing "REISSUE" or "REISSUED"
3. Select records with valid IREC_DOC_NBR (issued DL or ID number)

Enrique

From: Lee Guyette [mailto:LGuyette@sos.state.tx.us]
Sent: Friday, December 09, 2011 11:34 AM
To: Gomez, Enrique; Moon, Charles; John Mendoza
Cc: O'Connor, Ryan; Johnson, Lance
Subject: RE: jury wheel ftp address

Thanks again Enrique.

Hope all is well with the data processing. We did have a couple questions...

- Is the data from all of you DPS records or a subset?
- Other than filtering for "TEST", are there any other modifications being made to the data?

Lee Guyette, MSc PMP
Software Development Manager
Office of the Texas Secretary of State
(512) 423-5751
lguyette@sos.state.tx.us



From: Gomez, Enrique [mailto:Enrique.Gomez@dps.texas.gov]
Sent: Tuesday, December 06, 2011 10:21 AM
To: Lee Guyette; Moon, Charles; John Mendoza
Cc: O'Connor, Ryan; Johnson, Lance
Subject: RE: jury wheel ftp address

Lee,
I just got the go from our Chief of Staff. I will get my team to start working on it.

Enrique

From: Lee Guyette [mailto:LGuyette@sos.state.tx.us]
Sent: Tuesday, December 06, 2011 9:05 AM
To: Moon, Charles; Gomez, Enrique; John Mendoza
Cc: O'Connor, Ryan; Johnson, Lance
Subject: RE: jury wheel ftp address

Howdy Folks,

Has approval come your way yet? I understand there was another meeting on Friday and the green light was given.

Once again, the only filter we require is for removal of "test" data.

Lee Guyette, MSc PMP
Software Development Manager
Office of the Texas Secretary of State
(512) 423-5751
lguyette@sos.state.tx.us



From: Moon, Charles [mailto:Charles.Moon@dps.texas.gov]
Sent: Wednesday, November 30, 2011 9:08 AM
To: Gomez, Enrique; John Mendoza
Cc: O'Connor, Ryan; Lee Guyette; Johnson, Lance
Subject: RE: jury wheel ftp address

That was an old email before I got the direction from you all. I am waiting for the approval.

Charles Moon
512-590-5443 (m)

From: Gomez, Enrique
Sent: Wednesday, November 30, 2011 8:23 AM
To: Moon, Charles; John Mendoza
Cc: O'Connor, Ryan; Lee Guyette; Johnson, Lance
Subject: RE: jury wheel ftp address

Who is the ok coming from?. I have not received it yet.

Enrique

From: Moon, Charles
Sent: Tuesday, November 29, 2011 9:32 AM
To: John Mendoza
Cc: O'Connor, Ryan; Gomez, Enrique; Lee Guyette; Johnson, Lance
Subject: RE: jury wheel ftp address

I guess I missed that meeting. Sorry.

I can do a full run. It should be around 25M records.

Charles

From: John Mendoza [JMendoza@sos.state.tx.us]
Sent: Tuesday, November 29, 2011 9:26 AM
To: Moon, Charles
Cc: O'Connor, Ryan; Gomez, Enrique; Lee Guyette; Johnson, Lance
Subject: RE: jury wheel ftp address

Charles,

There was a meeting yesterday with folks from DPS. Enrique Gomez was there. DPS was asked for a complete file, no filters, except one. The removal of any "test" data. You may want to speak to Enrique about this. I believe y'all were waiting for today for an OK from someone to proceed.

Last time you created a self-extracting encrypted file which was compressed. I think the FTP site was able to handle it fine.

JohnM.

From: Moon, Charles [mailto:Charles.Moon@dps.texas.gov]
Sent: Tuesday, November 29, 2011 9:05 AM
To: John Mendoza
Cc: O'Connor, Ryan; Gomez, Enrique; Lee Guyette; Johnson, Lance
Subject: RE: jury wheel ftp address

John, sorry for the delay but I am only in the office one day each week during Dec. But I do check my email daily.

I can do a full run for you today. Please note that the file will be very large and I am not sure your FTP site will be able to handle it. I can do a run this morning if the remote access is available. Otherwise, I will have to go into the office this afternoon. Either way, I will take care of it.

More to come...

Charles

From: John Mendoza [JMendoza@sos.state.tx.us]
Sent: Monday, November 28, 2011 10:55 AM
To: Moon, Charles

Cc: O'Connor, Ryan; Gomez, Enrique; Lee Guyette
Subject: RE: jury wheel ftp address

Charles,

Can you send us another complete file? This time with no filters. I'm assuming the record count would be the 25 million mentioned in the email below. I think this may help us answer many of the questions we are getting.

We may have some questions on how you normally filter out "test" data that may be in the file.

Let me know if this would be possible and how long this could take for you.
JohnM.

From: Moon, Charles [mailto:Charles.Moon@dps.texas.gov]
Sent: Wednesday, November 23, 2011 11:02 AM
To: Gomez, Enrique; Lee Guyette; John Mendoza
Cc: O'Connor, Ryan
Subject: RE: jury wheel ftp address

Enrique, I just ran the counts again.

We have about 25M rows of total person records

- Running the request filters (all but SSN validation), we get about 22M records
- Removing duplicates using SSN + DPB combination, we get to 18.6M records

If SOS does not require valid SSNs, I can remove that filter and the count goes back close to 19M. However, we may introduce some duplicates since SSN is used in duplicate removal process. Please let me know which version is appropriate.

Thanks,

Charles Moon
512-590-5443 (m)

From: Gomez, Enrique
Sent: Wednesday, November 23, 2011 10:44 AM
To: Moon, Charles; 'LGuyette@sos.state.tx.us'; 'JMendoza@sos.state.tx.us'
Cc: O'Connor, Ryan
Subject: Re: jury wheel ftp address

Lee,
How many people do you think have Id's or dl's based on demographics?

Charles,
How many records do we have?

E

From: Moon, Charles
Sent: Wednesday, November 23, 2011 10:07 AM
To: Lee Guyette <LGuyette@sos.state.tx.us>; John Mendoza <JMendoza@sos.state.tx.us>

Cc: Gomez, Enrique; O'Connor, Ryan
Subject: RE: jury wheel ftp address

I agree, it appears to be. However, per your request, filters were used to remove duplicate, non-valid SSN, non-valid county, younger than 17yr 10mo, last name "Reissued" or "TEST", and no ID/DL number assigned records.

Did you want the unfiltered records?

Charles Moon
512-590-5443 (m)

From: Lee Guyette [mailto:LGuyette@sos.state.tx.us]
Sent: Tuesday, November 22, 2011 2:42 PM
To: Moon, Charles; John Mendoza; O'Connor, Ryan
Subject: RE: jury wheel ftp address

Charles and Ryan,

Thanks again for your help and prompt responses.

To be honest though, I am still struggling to understand the 14M number. In a state of 25+M, I would have assumed more than 56% of Texans would have drivers license or ID.

Am I missing something here?

Lee Guyette, MSc PMP
Software Development Manager
Office of the Texas Secretary of State
(512) 423-5751
lguyette@sos.state.tx.us



From: Moon, Charles [mailto:Charles.Moon@dps.texas.gov]
Sent: Tuesday, November 15, 2011 2:02 PM
To: John Mendoza; O'Connor, Ryan
Cc: Lee Guyette
Subject: RE: jury wheel ftp address

Potentially dangerous attachments were identified and removed from this message.

This message contains encrypted attachments that could not be verified virus free. Open only if you were expecting this message.

Sorry, try again... Change the type to exe.

The requirements was to remove all duplicates. There were 16 million IDs from which I removed duplicate SSN + DOB. We ended up with 14 million.

Charles Moon
512-590-5443 (m)

From: John Mendoza [mailto:JMendoza@sos.state.tx.us]
Sent: Tuesday, November 15, 2011 7:56 AM
To: Moon, Charles; O'Connor, Ryan
Cc: Lee Guyette
Subject: RE: jury wheel ftp address

Charles,

Our email system blocks any executable attachments. Please try zipping the EXE and resend the file or send via FTP as before.

Also, I'm concerned over the number of records. The annual jury wheel file (statewide driver/id file) sent late September/early November of this year had over 16 million records (this also filtered felons). The file you've sent me for this project is just over 14 million (which should still include the felons), so I was expecting more records than the jury file. Did we really remove 2 million records worth of duplicates? Or were those also due to some of the other filters?

I thought the people without SSN may have been those missing records, but if you are saying it is only 4,227 records, then I still need to ask if the record counts are correct.

Can you shed some light on accounting for the 2 million records shortfall?

JohnM.

From: Moon, Charles [mailto:Charles.Moon@dps.texas.gov]
Sent: Monday, November 14, 2011 1:34 PM
To: John Mendoza; O'Connor, Ryan
Cc: Lee Guyette
Subject: RE: jury wheel ftp address

This message contains encrypted attachments that could not be verified virus free. Open only if you were expecting this message.

John,

Same password and format as before. There should be 4227 records with blank SSNs.

Charles Moon
512-590-5443 (m)

From: John Mendoza [mailto:JMendoza@sos.state.tx.us]
Sent: Monday, November 14, 2011 1:26 PM
To: Moon, Charles; O'Connor, Ryan
Cc: Lee Guyette
Subject: RE: jury wheel ftp address

See responses in red below.

JohnM.

From: Moon, Charles [mailto:Charles.Moon@dps.texas.gov]
Sent: Monday, November 14, 2011 1:06 PM
To: John Mendoza; O'Connor, Ryan

Cc: Lee Guyette
Subject: RE: jury wheel ftp address

John, I need some clarification...

1. You want the remaining list of people who does not have SSN. Is that correct? If yes, I can get it to you today. Yes. Please send the remaining list of people that did not have SSN. These were inadvertently filtered during the filtering of those records that had the same Last Name, First Name, Date of Birth, and SSN.
2. I opened your sample file.
 - a. DPS – DL transactional system uses 11 digit numeric data type as system IDs. That is the reason why you have IDs with different number of digits. Adding leading "0" is acceptable since the system will ignore all leading zeros as they are not relevant for numeric data. OK, I'll add the leading zeroes, in the past we considered TDL/iD character fields and treated leading zeroes as significant digits.
 - b. Data without ethnicity should remain as they are as well as the people without counties. No problem.
 - c. You can predict their counties using zip code. But that would only be guessing. I can send you their zip codes, if you prefer. No need.

Charles Moon
512-590-5443 (m)

From: John Mendoza [mailto:JMendoza@sos.state.tx.us]
Sent: Monday, November 14, 2011 11:58 AM
To: Moon, Charles; O'Connor, Ryan
Cc: Lee Guyette
Subject: RE: jury wheel ftp address

FYI,

First, we'll need those missing records. Those that don't have an SSN. I realize you cannot check for duplicates in this set, but we still need those records for this project. if you can create and FTP that file, when would it be ready?

Second, I've attached a zipped file, which once unzipped, you'll find an encrypted exe file, which you can decrypt by using your same encryption password/passphrase. This file is the sample of the data related to the other questions. Ryan mentioned if the TDL/iD had 7-digits we can add the leading zero, but there were quite a few records that ranged from 1 digit all the way up to 7 digits. Would we treat all these as needing the leading zeroes?

Note: The first question is the most import part. The other questions, I can work around, I just wanted to get some confirmation that the records were valid records and that I can use them. Also, if adding the leading zeroes was the correct approach on my part.

Thanks.
JohnM.

Attorney-Client Privileged

TX_00055966

Attorney-Client Privileged

Attorney-Client Privileged

From: Moon, Charles [mailto:Charles.Moon@dps.texas.gov]
Sent: Tuesday, November 08, 2011 3:38 PM
To: John Mendoza
Cc: O'Connor, Ryan; Lee Guyette
Subject: RE: jury wheel ftp address

John, just let me know if you want me to make any changes to the data. We run SQL Server and CSV file format as our standard extract. Anything other than that, I can export using scripts.

I am out of office starting Thu until Monday of next week. So if you need something, you will have to catch me before 4PM tomorrow (Wed).

Thanks,

Charles Moon
512-590-5443 (m)

From: John Mendoza [mailto:JMendoza@sos.state.tx.us]
Sent: Tuesday, November 08, 2011 3:30 PM
To: Moon, Charles
Cc: O'Connor, Ryan; Lee Guyette
Subject: RE: jury wheel ftp address

FYI,

I believe I'm closer. I was able to filter out the non-printable characters and I'm able to read the file better now.

I'll try to load the file into our database. I wasn't expecting the date formats, but I'm hoping I can handle them too.

Will keep you updated.

JohnM.

From: Moon, Charles [mailto:Charles.Moon@dps.texas.gov]
Sent: Monday, November 07, 2011 4:28 PM
To: John Mendoza
Cc: O'Connor, Ryan; Lee Guyette
Subject: RE: jury wheel ftp address

John, this looks like code page issue.

All data sets we generate from the DPS warehouse are in Unicode (NVARCHAR). You may want to check to make sure you are opening it as Unicode text document.

Charles Moon
512-590-5443 (m)

From: John Mendoza [mailto:JMendoza@sos.state.tx.us]
Sent: Monday, November 07, 2011 4:05 PM
To: Moon, Charles
Cc: O'Connor, Ryan; Lee Guyette
Subject: RE: jury wheel ftp address

The record looks something like this:

ÿþ101ROSINSKI BENJAMIN ANTHONY

I'll try to filter the bad characters out. Will let you know.

JohnM.

P.S. I'm coming in a noon tomorrow, since I'll be working late due to the November Election.

From: Moon, Charles [mailto:Charles.Moon@dps.texas.gov]
Sent: Monday, November 07, 2011 3:58 PM

To: John Mendoza
Cc: O'Connor, Ryan; Lee Guyette
Subject: RE: jury wheel ftp address

Also, you should have 14,155,516 rows of data.

Charles Moon
 512-590-5443 (m)

From: Moon, Charles
Sent: Monday, November 07, 2011 3:44 PM
To: 'John Mendoza'
Cc: O'Connor, Ryan; Lee Guyette
Subject: RE: jury wheel ftp address

No, I did not include the header file but sent it to you in a separate email. The first row should look like this.

101ROSINSKI... BENJAMIN... ANTHONY.....

Is that what you have?

You can parse it using this position table.

Field Name	Data Type (size)	Position
[County Code]	[nvarchar] (3)	1-3
[Last Name]	[nvarchar] (40)	4-43
[First Name]	[nvarchar] (40)	44-83
[Middle Name]	[nvarchar] (40)	84-123
[Name Suffix]	[nvarchar] (10)	124-133
[SSN]	[nvarchar] (9)	134-142
[Gender]	[nvarchar] (1)	143
[TDL or TID Number]	[nvarchar] (11)	144-154
[Issue Date]	[date]	155-171
[Date of Birth]	[date]	172-188
[Ethnicity]	[nvarchar] (1)	189

Charles Moon
 512-590-5443 (m)

From: John Mendoza [mailto:JMendoza@sos.state.tx.us]
Sent: Monday, November 07, 2011 3:08 PM
To: Moon, Charles
Cc: O'Connor, Ryan; Lee Guyette
Subject: RE: jury wheel ftp address

Yes. I was able to decrypt the file.

However, there's something strange with the beginning of the file. I'm getting some special characters at the start of the file, followed by the data. Possibly it's the column header record you mentioned you were going to add at the top. I don't see the column headers, just the special characters.

It's giving me some trouble reading the file.

JohnM.

From: Moon, Charles [mailto:Charles.Moon@dps.texas.gov]
Sent: Monday, November 07, 2011 2:16 PM
To: John Mendoza
Cc: O'Connor, Ryan
Subject: RE: jury wheel ftp address

John, were you able to run it to extract the text file?

Charles Moon
512-590-5443 (m)

From: John Mendoza [mailto:JMendoza@sos.state.tx.us]
Sent: Monday, November 07, 2011 1:47 PM
To: Madipadga, Vanisree
Cc: Moon, Charles; Lee Guyette
Subject: RE: jury wheel ftp address

Yes, I have received it.

JohnM.

From: Madipadga, Vanisree [mailto:Vanisree.Madipadga@dps.texas.gov]
Sent: Monday, November 07, 2011 1:33 PM
To: John Mendoza
Cc: Moon, Charles
Subject: FW: jury wheel ftp address

Hi,
The file SOS_2f.exe transferred to SOS and I can see the file on their server.

Thanks,
Vani

From: Moon, Charles
Sent: Monday, November 07, 2011 1:11 PM
To: Madipadga, Vanisree
Subject: RE: jury wheel ftp address

Done.

Charles Moon
512-590-5443 (m)